DOCUMENT RESUME

ED 393 828 SP 036 573

AUTHOR Nweke, Winifred; Noland, Juanie
TITLE Diversity in Teacher Assessment: What's Working, What's Not?
PUB DATE Feb 96
NOTE 15p.; Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education (48th, Chicago, IL, February 21-24, 1996).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Black Colleges; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; Grade Point Average; Higher Education; *Minority Group Teachers; *Performance Based Assessment; *Portfolio Assessment; Preservice Teacher Education; *Student Evaluation; Teacher Certification; *Teacher Education Programs
IDENTIFIERS *Tuskegee University AL

ABSTRACT

The purpose of this study was to evaluate the effectiveness of using performance and portfolio assessment techniques to diversify assessment in a minority teacher education program at Tuskegee University (Alabama). Both performance and portfolio assessments served as the exit examination. These techniques were used for three reasons: a state traditional certification examination was thrown out by the Alabama courts because it had negative impact on minority students; the university serves a predominantly minority population; and traditional tests have not been very valid measures of performance tasks such as teaching. Grades from portfolio and performance assessments were obtained from the files of 30 graduates of the teacher education program. These were correlated with grades in methods courses, foundation courses, and grade point average (GPA). Significant correlations were found between methods classes, on the one hand, and performance measures and portfolios on the other. Though the performance assessment measure was significantly correlated to GPA, no relationship was found between portfolio and GPA. Also, the portfolio score was not significantly correlated to the performance measure. The results suggested that performance and portfolio measures elicit skills and abilities that are relatively independent of those elicited by traditional tests as represented by the GPA and are independent of each other. It was concluded that though performance and portfolio measures show good promise for diversifying assessment for teacher certification, portfolio assessment itself needs to be improved. The appendix provides portfolio criteria from the Educational Testing Service Praxis Series of Assessments. (Contains 24 references.) (ND)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 393 828

Running head: Diversity in Teacher Assessment

Diversity in Teacher Assessment:

What's Working, What's Not?

Winifred Nweke and Juanie Noland

Tuskegee University

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

W Nweke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Paper presented at AACTE
1996 Annual Meeting
February 21-24, 1996, Chicago, Illinois

SP036573

BEST COPY AVAILABLE

Abstract

The purpose of this study was to evaluate the effectiveness of using performance and portfolio assessment techniques to diversify assessment in a minority teacher education program. Performance and portfolio assessments served as our exit examination. These techniques were used for three reasons: a state traditional certification examination was thrown out by the Alabama courts because it had negative impact on minority students; we serve a predominantly minority population; and teaching is a performance task and traditional tests have not been very valid measures of performance tasks such as teaching. Data were obtained from the files of 30 graduates of the teacher education program. Students' grades from portfolio and performance assessments were correlated with grades in methods courses, foundation courses, and GPA. There were significant correlations between methods classes, on the one hand, and performance measure (r (30) = 0.41, $p < .05$) and portfolio (r (30) = 0.32, $p < .05$), on the other. Though the performance assessment measure was significantly correlated to GPA (r (30) = 0.40), there was no relationship between portfolio and GPA (r (30) = 0.09). Also, portfolio score was not significantly correlated to the performance measure. The results suggested that performance and portfolio measures elicit skills and abilities that are relatively independent of those elicited by traditional tests as represented by the GPA and are independent of each other. Inter-rater reliability coefficients on portfolio assessment, based on individual Praxis III Series criteria, ranged from r_{xx} (25) = 0.0 to $r_{xx}(25)$ = 0.38. It was concluded that though performance and portfolio measures show good promise for diversifying assessment for teacher certification, portfolio assessment itself needs to be improved. Suggestions for improvement were made.

Diversity in Teacher Assessment:
What's Working, What's Not?

The purpose of this study was to determine to what extent portfolio assessment has helped achieve the goal of assessment diversification in a teacher education exit examination system.

Perspectives and Theoretical Framework

Teacher evaluation based on standardized one-shot objective examination is now widely recognized as insensitive to a host of diverse and important teacher education attributes and contextual variables. Teaching is recognized as complex, holistic, highly integrated and contextualized. (Athanases, 1990; Barton & Collins, 1993; Dwyer, 1993). Thus, standardized pencil-and-paper tests appear not to be appropriate, authentic, or valid enough to use for decisions regarding teaching performance, certification and promotion (Chittenden, 1991; Stiggins, 1986; Wiggins, 1989). In addition to being inappropriate for use with teaching, standardized teacher examinations tend to have negative impact on minorities (Bredekamp & Shepard, 1989).

The argument against traditional approaches to teacher assessment was strengthened in recent years by the courts' rejection of many state-developed teacher certification examination. The failure rate of blacks and other minorities on such examinations, in Alabama, was disproportionate to that of whites. Furthermore, these examinations often are said to lack criterion-related validity (Bredekamp & Shepard, 1989). Other authors dealing with younger populations have made similar observations about traditional tests in general. For example, Richert and McDonnel (1982) in the National Report on Identification suggest that the use of paper-and-pencil tests leads to the exclusion of minorities from gifted programs. Also, Shaklee (1992) suggested that performance-based assessments are valid approaches for documenting the potential of under-represented groups in gifted programs. According to Chittenden (1991), authentic assessments may more accurately reflect holistic approaches to teaching than traditional ones. Thus, given the population we serve (which is predominantly black), and the purpose of testing

(which is demonstration of teaching competence) it made sense to use a performance-type test for our exit examintion, rather than a traditional test.

In an attempt to provide more varied, valid and authentic measures of teaching competence, researchers and practitioners have revived interest in old techniques long ignored, and developed new ones, including: journals, portfolios, and various forms of performance-based methods such as classroom observations. Use of these techniques in evaluating teaching competence has spawned hundreds of journal articles and books on performance assesment in general and portfolio assesment in particular. Regardless of their increasing popularity, however, many of the techniques have serious weaknesses in objectivity, reliability, validity and generalizability (Linn, 1993.)

Consistent with a national trend toward performance-based assessment, Educational Testing Service (ETS), developers of the widely-used National Teachers Examination (NTE), developed the Praxis Series III: Classroom Performance Assessments which are based on several years of research on the conceptualization of teaching as a highly integrative, productive, and complex activity (Dwyer, 1993). Teaching entails engaging students as active learners to induce changes in their previous knowledge — a constructivist view of learning. In the constructivist theory tradition, learning is viewed as an active process of constructing meaning. Thus, teacher assessments should reflect a diversity of classroom contexts (including differences in subject matter, students' backgrounds and styles of learing). Given the complexity of this approach to teacher assessment, we need trained, sensitive evaluators. One example of such an approach is found in the Praxis III: Classroom Performance Assessments, whose assessors must participate in five days of training and pass a proficiency test before certification of proficiency by ETS.

ETS is not alone among teacher evaluation organizations in their development of criteria for demonstrating quality teaching. Other groups include The National Council for the Accreditation of Teacher Education (NCATE), the National Board for Professional Teaching Standards, the National Policy Board for Educational Administration (work led by practicing educators), the Interstate New Teacher Assessment and Support Consortium (work led by the nation's chief state school officers, and many state departments of education, include performance criteria and standards in their assesment of teachers. Many states have also used or are using performance assesment in grades K-12, for example: Vermont (Koretz, et al, 1994); Connecticut; California, and New York (Baron, 1991). Some benefits reported in these sites include more equitable evaluation of minorities than with traditional tests, divergent

thinking, cooperative learning, creativity, and interpendence among team members. Performance tests also have been reported to facilitate equal learning among boys and girls and among various ethnic groups. In small groups, minorities who normally remain segregated in regular classrooms cease to be minority, according to Mangani (as cited in Baron, 1991).

Portfolio assessment often is included in performance-based assessment programs. Portfolios are used for various purposes. Because of portfolios' face validity (Barton & Collins, 1993), many educators use them as substitutes for traditional tests (Black, 1990; Dagvarian, 1989; Woodrow, 1982), while others use them as supplements (Bird, 1990; Forrest, 1990; Nweke, 1990; O'Brien, 1990; Wolf, 1990). Portfolio technique is also used to assess general education (Hunter, 1989), and as supplementary resume' (Nweke, 1990).

How Portfolio Assessment is Used at Tuskegee University to Address Preservice Teacher Diversity: Methods

Since 1993, teacher education programs in Alabama are required by the State Department of Education to develop their own exit examinations and administer them to their students. This policy followed a lengthy legal battle regarding the state's previous state-developed pencil-and-paper certification examination which, according to the courts, had a negative impact on blacks.

In developing our exit examination, Tuskegee University decided that simply constructing another pencil-and-paper test would not adequately address diversity issues, especially in view of our predominantly-black student body. Other reasons for our decision were the controversy surrounding the state's pencil-and-paper test of teacher competency, the large number of pencil-and-paper tests which students take in their college courses, and the emerging data regarding performance-based and portfolio assessment. Our solution was what we call the Comprehensive Examination. The Comprehensive Examination is in two parts: Part 1 is a performance-based evaluation. Part 2 is a portfolio. Parts 1 and 2 use the same Praxis Series criteria for evaluation (See Appendix). The rating form for Part 1, the performance-based evaluation, is completed collaboratively by a student's cooperating teacher and university supervisor, at the end of a student's internship, and is designed to assess the student's classroom performance and ability to integrate theory and practice. Students' ratings are based on their performance during student teaching over a period of ten weeks.

The Portfolio, Part 2, is developed by student teachers to demonstrate that each Praxis criterion is met. The portfolio assessment allows students to show, in diverse ways, that they have met each criterion and that they can integrate theory and practice. Each portfolio is evaluated by two to four readers, comprising the students' cooperating teachers, university supervisors, and two other Tuskegee University professors. A portfolio grade is the average of scores from all raters. The portfolio and performance test technically are more in accord, than traditional tests, with our newly-developed and implemented "Constructivist Reflective" model in the School of Education. In this model, we encourage personal and reflective responses from students, through journal writing, personal histories, and reflection on lessons taught. We chose the portfolio and performance techniques for the exit exam because, for our purposes, they better reveal a student's development toward becoming a teacher and because, at least conceptually, they seem to have higher criterion-related validity than more traditional tests. Such methods also recognize that attainment of a performance outcome —for example, "building instruction on students' prior learning and academic strengths" — can best be verified through observation or performance assessment. In the next section the method for the study is described.

Data Source

Data were obtained from the files of 30 secondary, elementary and early childhood students who graduated from Tuskegee University between 1993 and 1995. Specifically, students' grades on portfolios, performance assessments, GPAs, methods, and foundations courses were obtained. Portfolio and classroom performance grades were correlated with students' overall GPA, and average grades on foundations and methods courses. The GPA was computed on all courses except for student teaching and portfolios. Interrater reliability among cooperating teachers and teacher education faculty was also investigated.

Results

The classroom performance assessment showed a positive and significant correlation with overall grade point average (GPA) ($r(30) = 0.399$, $p < .05$) and with methods courses ($r(30) = 0.413$, $p < .05$). It showed no significant relationship, however, with portfolio measure ($r(30) = .219$, $p > .05$) and foundation courses ($r(30) = 0.162$), $p > .05$). (See Table 1).

The portfolio measure showed a significant relationship with methods classes ($r(30) = 0.39$, $p < .05$). However it had no relationship with GPA ($r(30) = 0.087$, $p > .05$), and a positive but nonsigificant relationship with foundation courses ($r(30) = .150$, $p > .05$ and the Classroom Performance Assesment ($r(30) = .219$, $p > .05$).

It was also found that the inter-rater reliability coefficients between cooperating teachers and teacher education faculty were very low. As shown in Table 2, these coefficients range from a low of $r_{xx} = 0.0$ to a high of $r_{xx} = 0.38$, with a median reliability coefficient of $r_{xx} = 0.09$. All inter-rater reliability figures between cooperating teachers and faculty raters were zero or negative, except one. Inter-rater correlations were higher among the teacher education faculty. The lowest was $r_{xx} = .02$ and the highest was $r_{xx} = 0.81$, with a median of $r_{xx} = 0.31$.

Discussion

These results suggest that portfolio assessment elicits different skills and abilities from those measured by traditional tests. Therefore, portfolios show promise for diversifying measures used in teacher certification decisions.

Interestingly, the one positive correlation between cooperating teachers and faculty raters is between the faculty member whose ratings correlated negatively with most of the other faculty raters.

The low inter-rater coefficients are a cause for concern. The first question they raise is: Why are the correlations so low? Low to medium inter-rater reliability coefficients are not uncommon in the area of portfolio assessment (Koretz, et al, 1992; Nystrand et al, 1993). An average reliability figure of 0.43 was reported on separate areas of writing portfolios in the Vermont portfolio assessment program (Koretz et al, 1992). The average reliability increased to 0.58 when computed on total scores. There are some possible explanations for the low reliability coefficients. The first is generosity or leniency error. There were indications that cooperating teachers were trying to be nice and to "help out" their student teachers. In these cases, the verbal and written comments from the cooperating teachers did not match the perfect or near-perfect grades they awarded the student interns.

There were evidences of inadequate comprehension of some criteria. For example, Criterion 1.2 under Content Knowledge for Teaching appeared to have been misunderstood. It states that "teaching intern demonstrates an understanding of the connections between the content of an instructional event and

what was studied previously or remains to be studied in the future." The criterion is really asking for evidence that the student teacher understands the scope and sequence of content taught. Entries under Criterion 1.2 were not always satisfactory.

A third explanation for low inter-rater reliability coefficients is restriction of range. The student interns are a highly homogeneous group due to prior selection. To qualify for student internship students would have taken and passed all required courses with a GPA of 2.5 or better. They also had to have peformed satisfactorioly at an interview. Thus, it would not be unusual for the interns to make similar grades on the performance and portfolio assessments. Also, the reliability figures were based on ratings on a scale of 1-5 assigned to individual criteria rather than total portfolio score. For example, a faculty member awarded a score of 94 to a Portfolio, while the cooperating teacher awarded 95, and yet the inter-rater reliability between the two raters was 0.09!

The final explanation is lack of training. Neither the teacher education faculty nor cooperating teachers received any formal training in portfolio or performance assessment. Training will be conducted.

Educational Importance of Study

More and more, teacher education programs use portfolios as part of their assessment systems. Teacher educators are seeking understanding of the role of portfolios in teacher education and the relationship of portfolios to more traditional assessment techniques. Educators especially want to know what specific skills and abilities portfolios measure, how valid they are as a measure of teacher competence, and their adequacy in addressing the issue of diversity. This study adds to the growing knowledge base of portfolio assessment in teacher education regarding the relationship of portfolios to other evaluation measures, and common problems in implementing portfolio assessment.

References

Athanases, S.Z. (1990). Using teacher portfolios to assess planning, instruction, and teacher reflection in elementary language arts. Paper presented at the American Educational Research Association Annual Meeting, Boston, MA.

Baron, J.B. (1990). Performance assessment: blurring the edges of assessment, curriculum, and instruction. In G. Kuim & S. M. Malcolm (eds.). Science assessment in the service of reform (pp. 247-266). Washington, DC: American Association for the Advancement of Science.

Barton, J. and Collins, A. (1993). Portfolios in teacher education. Journal of Teacher Education, 44 (3), 200-210.

Bird, T. (1990). The school teacher's portfolio: An essay in possibilities. In J. Millman and L. Darling-Hammond (Eds.), Handbook of teacher evaluation: Elementary and secondary personel. Beverly Hills: Sage.

Bredekamp, S. & Shepard, L. (1989). How best to protect children from inappropriate school expectations, practices and policies. Young Children, 44 (3), 14-24.

Chittenden, E. (1991). Authentic assessment, evaluation and documentation of student performance. In V. Perrone (ed.), Expanding Student Assessment, (pp.21-33). Alexandria, VA: Association for Supervision and Curriculum Development.

Dagvarian, D. (1989). Portfolio assessment. ERIC ED306894.

Dwyer, C.A. (1993). Teaching and diversity: Meeting the Challenge for Innovative teaching. Journal of Teacher Education, 44 (2), 119-129.

Forrest, A. et al. (1990). Portfolio-assisted assessment of general education . (Draft). Presented at the annual conference of the American Association of Higher Education, Washinton, D.C.

Gollnick, D.M. (1994). New accreditation standards: Focus on performance. NCATE Quality Teaching, 4(1), 6-7.

Hunter, S. (1989). Evergreen State College. In P. Hutchings (1989). Portfolio method of assessment: A sampling of campus practice. Washington, D.C.: American Association of Higher Education.

Linn, R.L. (1993). Educational assessment: Expanded expectations and challenges. Educational Evaluation and Policy Analysis, 15 (1), 1-16.

Moss, P.A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. Review of Educational Research, 62, 229-258.

National Council for Teachers of Mathematics (1992). Standard 3: Appropriate assessment methods and uses. NCTM Evaluation Standards. Author.

Nweke, W.C. (1990). Instructions for portfolio preparation. Document prepared for the Alabama Consortium for Minority Teachers.

O'Brien, K. (1990). Portfolio assessment at Alverno College. Paper presented at the AAHE conference, June, 1990.

Praxis Series of Assessments. ETS Policy Notes, 3 (2), Spring 1991.

Shaklee, B. (1992). Identification ofyoung gifted children. Journal for Education of the Gifted, 15 (2), 134-144.

Shavelson, R.J., Baxter, G.P. and Gao, X. (1993). Sampling variability of performance assessment. Journal of Educational Measurement, 30 (3), 215-232.

Stiggins, R.J., Conklin, M.F., Bridgeford, J.J. Classroom assessment: a key to effective education. Educational Assessment: Issues and Practice, 5 (2), 15-17.

Viechnicki, K.J. Barbour, N., Shaklee, B., Rohrer, J., and Ambrose, R. (1993). The impact of portfolio assessment on teacher classroom activities. Journal of Teacher Education, 44(5), 371-377.

Wiggins, G. (1989). Teaching to the (authentic) test. Educational Leadership, 46 (7), 41-47.

Wise, A. (1995). NCATE's emphasis on performance: Presidential perspective. NCATE Quality Teaching, 4(3), 6-7.

Wolf, K. (1990). The school teacher's portfolio: Practical isues in design, implementation, and evaluation. Teacher Assessment Project, Stanford University, Stanford, CA.

Appendix

Portfolio Criteria[1]

1.0 CONTENT KNOWLEDGE FOR TEACHING

1.1 Professional teaching intern demonstrates knowledge of content through instructional events that are logically sequenced and that are sound and accurate reflections of the content.

1.2 Professional teaching intern demonstrates an understanding of the connections between the content of an instructional event and what was studied previously or remains to be studied in the future.

1.3 Professional teaching intern can create or select curricular materials and other resources, learning activities, and evaluation strategies that are clearly linked to the intent or goal of the instructional event.

2.0 TEACHING FOR STUDENT LEARNING

2.1 Professional teaching intern builds instruction on students' prior learning and academic strengths.

2.2 Professional teaching intern accommodates students' individual interests, developmental levels, and cultural resources by engaging them in a variety of learning activities.

2.3 Professional teaching intern monitors students' understanding of content through a variety of means, providing feedback to students to assist learning, and adjusting learning activities as the situation demands.

2.4 Professional teaching intern makes expectations clear to students, setting high expectations for all, and helping students take responsibility for their own learning.

2.4 Professional teaching intern makes content comprehensible to students through clear and focused explanations, and meaningful examples, analogies, metaphors, and/or demonstrations.

2.5 Professional teaching intern encourages students to extend their thinking beyond factual knowledge.

3.0 CLASSROOM COMMUNITY FOR STUDENT LEARNING

3.1 To the extent possible in this classroom, professional teaching intern creates a purposeful and well-functioning learning community with convenient and well-understood classroom routines.

3.2 To the extent possible in this classroom, professional teaching intern creates an attractive and safe physical environment arranged in ways conducive to student learning.

---

[1] The criteria are from Educational Testing Service's Praxis Series of assessments. The ETS criteria were published in *ETS Policy Notes*, Volume 3, Number 2, Spring 1991.

3.3 To the extent possible in this classroom, professional teaching intern makes standards of behavior and consequences of misbehavior clear to students, and handles disruptions efficiently and with respect.

3.4 To the extent possible in this classroom, professional teaching intern creates a classroom climate that ensures equity and respect for and among students.

3.5 To the extent possible in this classroom, professional teaching intern establishes and maintains rapport with students.

3.6 To the extent possible in this classroom, professional teaching intern communicates high expectations for the learning and behavior of all students.

4.0 TEACHER PROFESSIONALISM

4.1 The professional teaching intern is able to reflect on and analyze his/her own instruction; characterize successes and failures; identify actions taken and rationales for them; and determine the extent to which instructional goals are met.

4.2 Professional teaching intern is able to explain how insights gained from instructional experiences can be used to improve instruction.

4.3 Professional teaching intern demonstrates personal responsibility for student learning.

4.4 To the extent possible in this setting, professional teaching intern is able to build professional relationships with colleagues to share teaching insights and coordinate learning activities for students.

4.5 To the extent possible in this setting, professional teaching intern is able to communicate with parents regarding student learning, and,where appropriate, interact effectively with the community.

Table 1

Correlation Among Traditional and Non-Traditional Measures

	GPA	FDN	Methods	STTCHG
Grade Point Average (GPA				
Foundation (FDN)	0.738*			
Methods Courses	0.666*	0.655*		
Student Teaching (STTCHG)	0.399*	0.162	0.413*	
Portfolio	0.087	0.150	0.320*	0.219

n = 30, * = significant at 0.05 level

Table 2

Inter-Rater Reliability Coefficients

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
Rater 1										
Rater 2			.31							
Rater 3										
Rater 4										
Rater 5	.13	.64	.31	.57						
Rater 6					.02 .76					
Rater 7	-.02 -.10				.56 .62 .18	.23				
Rater 8				.23	.32					
Rater 9		.61			.62 .81					
Rater 10	-.04				.77 .18 .45	.66	.31	.47	.57	
PSchl	.38				-.04 -.40	.00	-.19 -.08			

n = 25 items
range of scores per item is 1-5